How Sensational Content and Outgroup Cues Strengthen Support for Violence and Anti-Muslim Policies

Jeffrey Javed* Blake Miller†

Abstract

Existing research has acknowledged online information as a source of violent and discriminatory behavior. However, this research has primarily focused on its diffusion rather than its substantive effects. This study examines what kind of information drives violence and discrimination, testing the effects of sensationalization, outgroup cues, and public opinion perception on support for violence and anti-Muslim policies. We test this with an online survey experiment via a realistic, interactive website treatment detailing a homicide story in small-town America. We find that sensationalist language increased individuals' support for violence by provoking feelings of anger and fear while identifying the suspect in the homicide as a Muslim refugee, versus specifying no outgroup affiliation, increased support for anti-Muslim policies. Lastly, perceived public support for violence increased the likelihood of upvoting or writing violent comments. This study contributes to our understanding of the effects of sensational news and public debate on online content moderation.

*Research Fellow, Weiser Center for Emerging Democracies, University of Michigan (Email: jeffrey.javed@gmail.com)

†Assistant Professor, London School of Economics (Email: b.a.miller@lse.ac.uk)

1 Introduction

Social media companies have identified sensationalized content and false news as a significant threat to the integrity of their online communities across the globe (**?**Hern 2018; Lyons 2018). Media coverage has frequently linked the dissemination of false news stories—particularly those detailing sensationalized accounts of criminality or moral wrongdoing—to politically-motivated attempts to mobilize violence against vulnerable minority groups. In India, the largest market for Facebook's WhatsApp platform, political parties have used WhatsApp to spread false and sensationalized stories of murder and religious sacrilege by Muslims to inflame Muslim-Hindu tensions (Parth and Bengali 2018). A United Nations report claimed that the military in Myanmar had disseminated sensational, false news on Facebook that may have driven genocidal violence against the Rohingya (Mozur 2018). Non-state actors similarly use sensationalized and false news toward violent ends. The Sri Lankan government blocked Facebook after witnessing an increase in mob violence in response to fake, sensationalized content about the alleged misdeeds of Muslims spread by extremist groups (Goel, Kumar and Frenkel 2018). In the US, the proliferation of sensational anti-immigrant news from the alt-right has coincided with a dramatic rise in the number of reported hate crimes, as recorded by the FBI and the Southern Poverty Law Center (Barrett 2018). While systematic data on the intent of American alt-right content creators are unavailable, there is anecdotal evidence that some of these producers intentionally seek to promote or normalize violence. Notably, a leaked "style manual" from the neo-Nazi website *The Daily Stormer* instructed its writers: "It's illegal to promote violence on the Internet. At the same time, it's totally important to normalize the acceptance of violence as an eventuality/inevitability" (Feinberg 2017).

Do news stories featuring sensationalized claims of wrongdoings by minority groups affect mass attitudes toward the acceptability of violence and anti-minority policies? Despite growing concerns over the global proliferation of false news through social media, we know more about its patterns of diffusion than its effects (Brady et al. 2017; Lazer et al. 2018;

Vosoughi, Roy and Aral 2018); and we know nearly nothing about its effects on violent attitudes and behavior. Social media provides strong economic incentives to produce and spread sensationalized, outrage-inducing content (Crockett 2017). However, we do not know if sensational content increases the acceptability of violence, nor do we understand the cognitive mechanisms by which such an effect would occur. Research on related phenomena of hate speech and propaganda provide scant guidance: research has focused mainly on the legal and ethical dimensions of hate speech moderation and less on its real-world effects (Gates Jr et al. 1996; Waldron 2012), and observational research on genocidal propaganda has produced mixed results (Fujii 2004; Hagan and Rymond-Richmond 2008; Straus 2007; Yanagizawa-Drott 2014). While psychologists have found links between the consumption of violent media and aggressive behavior, this research has been restricted to children (Anderson and Sobel 2003; Drabman and Thomas 1974; Huesmann and Miller 1994).

Our study examines the effects of sensationalization, outgroup cues, and public opinion perception on support for violence and anti-Muslim policies. We used an online 2^3 factorial experiment with a realistic, interactive website treatment detailing a homicide story in small-town America, written in the style of a sensationalized, anti-immigrant news article that routinely appear on alt-right news platforms. Through an analysis of open-ended survey text, and online behavioral data, we find that sensational content—which we define as text, images, and videos that employ moral-emotional language[1] to present claims of moral violation or wrongdoing—increases individuals' endorsement of violence by provoking anger and fear. Also, we find that identifying the suspect in the homicide as a Muslim refugee, versus specifying no outgroup affiliation, increased support for anti-Muslim policies, while perceived public support for violence increased the likelihood of upvoting or writing violent comments.

[1]Moral emotions are feelings relating to "evaluations of societal norms" (Brady et al. 2017), specifically "feelings that stem from violating evaluative cultural codes, that is, codes that indicate what is good or bad or right or wrong in a society" (Stets 2012). To understand moral-emotional language, we draw on (Brady et al. 2017)'s dictionary of moral-emotional language, which contains words in the overlap between previously validated dictionaries of moral language and emotional language from the Linguistic Inquiry and Word Count (LIWC).

This argument builds on research in social psychology that suggests that exposure to moral violations lowers an individual's threshold for using violence against perceived norm violators by provoking outrage and providing justifications for their punishment (Baumeister 1999; Beck 1999; Crockett 2017; Fiske and Rai 2014; Goldberg, Lerner and Tetlock 1999) and explores the mechanisms by which political and non-state actors can mobilize a critical mass of support for violence among the public. The rest of this article proceeds as follows. The next section briefly reviews the literature on political violence, political mobilization, and emotion. We then detail our experimental design and variable measurement strategies and present our results. The final section concludes with a discussion of how these findings apply to the sensational news, the false news phenomenon, and violent political mobilization more generally.

2 Literature Review

The scholarship on collective violence, genocide, and war has long noted the outsized role of fabricated, sensational content in promoting violence against a targeted outgroup (Charny 2019; Cohn 1967; Dower 1986; Fein 1979; Herf 2006; Hill 1995; Goldhagen and Wohlgelernter 1997; Goldhagen 2009; Tsesis 2002). For example, the circulation of Cotton Mather's *Memorable Providences* detailing "real" accounts of bewitchings of innocent children is thought to have fueled the persecution of "witches" in the Massachusetts Bay Colony (Hill 1995); the *Protocols of the Elders of Zion*, which described a plan for Jewish world domination, was an important precursor to anti-Semitic Nazi mobilization (Cohn 1967; Herf 2006); and sensationalized accounts of Japanese atrocities against American POWs featured heavily in American anti-Japanese propaganda designed to recruit soldiers and bolster their fighting spirit (Dower 1986).

Few scholars, however, have directly tested the causal effects of content on violent attitudes, and the results of these studies have been mixed. Looking at political attitudes,

Kalmoe (2014) finds that violent rhetoric interacts with trait aggression to increase support for political violence. In the context of genocidal violence, Hagan and Rymond-Richmond (2008) find that the use of dehumanizing racial epithets during attacks on black African populations in Darfur correlated with more extreme violence. Other scholars have challenged the causal link between hateful content and violence. Looking at the Rwandan genocide, Straus (2007) argues that the spatial coverage of Hutu-controlled "hate radio" cannot explain the geographic variation in the onset of violence, while others have argued that anti-Tutsi propaganda was critical to driving violence or, at the very least, normalizing it (Fujii 2004; Yanagizawa-Drott 2014).

Moreover, it is unclear what kinds of content promote violence. Though scholars have emphasized the primacy of dehumanizing content in genocidal violence (Fein 1979; Charny et al. 1982), there is little empirical evidence for its efficacy. A major underexplored alternative is moral frameworks that justify the righteousness of violence (Viterna 2014). Social psychologists have long emphasized the significance of morality—conceptions of right and wrong traits and behaviors—in understanding participation in violence, private and political (Bandura, Underwood and Fromson 1975; Baumeister 1999; Beck 1999; Fiske and Rai 2014). Because using violence requires first overcoming formidable moral-emotional reservations, moral narratives that condone and justify violence can erode these otherwise powerful restraints on violent behavior (Baumeister 1999; Beck 1999; Fincher and Tetlock 2016). Moralistic content may also mobilize participation in violent causes by drawing upon individuals' latent moral convictions. Wood (2003) emphasizes the pleasure of agency that participants derive from acting on their moral convictions in a movement, which Viterna (2014) extends to the willingness of citizens to accept violence in the name of righteous causes, movements where "interested publics believe that the enactors of political violence are defending society's most vulnerable and protecting a morally legitimate social order." Indeed, Kirkpatrick (2008) documents a long history of extrajudicial and vigilante violence in America that invokes revolutionary values of freedom, justice, and democracy to frame itself as righteous.

Anger is an important emotional mechanism to consider when understanding the effects of sensational content on violent behavior. An abundant literature has analyzed the role of emotion in voter mobilization (Ansolabehere and Iyengar 1997; Banks 2014; Brader 2005, 2006; Freedman and Goldstein 1999; Huber et al. 2015; Marcus, Neuman and MacKuen 2000; Mendelberg 2001), and, more recently, in the context of misinformation (Vosoughi, Roy and Aral 2018); however, the emotional microfoundations of violent mobilization remain underexplored (Viterna 2013). Anger appears to be the most relevant emotion in mobilizing violent behavior. Unlike fear, which tends to demobilize, social psychologists have found that anger is a "mobilizing" emotion that increases political participation (Ansolabehere and Iyengar 1997; Banks 2014; Lerner and Keltner 2001; Ryan 2012; Valentino, Hutchings and White 2002; Valentino et al. 2011). Moreover, anger is a punitive emotion that motivates people to "shame and punish wrongdoers" perceived of having committed normative violations (Crockett 2017; Goldberg, Lerner and Tetlock 1999). Although research has shown that messages that provoke anger are more likely to mobilize political participation (Ryan 2012; Valentino, Hutchings and White 2002; Valentino et al. 2011) and moral-emotional content is more likely to be shared (Brady et al. 2017), it is unclear if such content can influence violent attitudes or engagement and whether anger mediates this relationship. Contextualizing this anger on social media is key for two reasons. First, sensational content rich in moral-emotional language is more likely to be shared or encountered online (Brady et al. 2017; Crockett 2017). Second, people report much higher levels of outrage when encountering this kind of content on social media than in-person or on traditional forms of media like TV and radio (Crockett 2017).

3 Theory and Hypotheses

We argue that content that sensationalizes transgressive behavior by utilizing highly moralized and emotional language makes individuals more punitive (sensational content

hypothesis). This type of sensational content is often found in false news content. We hypothesize that the mechanism linking sensationalized content and increased support for violence is outrage—i.e., anger that arises from perceived violations of cultural norms of right or wrong behavior (Crockett 2017; Stets 2012; Turner and Stets 2006) (outrage-aggression hypothesis). If there is a cue that explicitly links the norm violator to an outgroup, we predict that individuals will be more likely to support negative sanctions on that entire outgroup (outgroup cue hypothesis). We expect an interaction between content type and the outgroup cue: we predict that sensationalized content and the presence of an outgroup cue will greatly increase support for sanctions on the outgroup.

Also, we consider that individual support for violence will be partly a function of perceptions of peer attitudes (bandwagoning hypothesis). Perceived public opinion shapes attitudes and behavior (Neubaum and Kramer 2017; Noelle-Neumann 1993). Public opinion perception may be particularly relevant for punitiveness since a perceived audience to one's decision-making increases an individual's willingness to punish moral transgressions (Kurzban, DeScioli and O'Brien 2007) because proposing punishment of moral transgressors signals to ones peers that one is of high quality and potentially a good cooperator (Fessler and Haley 2003; Gintis, Smith and Bowles 2001). The perceived audience consensus on treating a violation may influence individual attitudes: if one's peers are uniformly signaling support for punishment, we expect that an individual will want to conform to the peer consensus to avoid reputational loss. We do not hypothesize any interaction effects between the peer effects treatment and the other treatments.

4 Experimental Design and Variable Measurement

Observed correlations between exposure to sensational content and the endorsement of violence could be due to selection effects, since people prone to violence might be attracted to such news in the first place. For this reason, we used a survey experiment embedded in

a realistic, fully interactive online news article about a local homicide in the US to examine how content sensationalization, outgroup cues, and violent comments by other online users may mobilize or inhibit violent attitudes and online engagement. We chose a homicide story for several reasons. Homicide stories have long been a mainstay of national and local news media in America and a possible source of anti-outgroup bias (Gilliam Jr and Iyengar 2000). Moreover, as mentioned earlier, there is plentiful anecdotal evidence that misleading crime stories feature heavily in false news content used to mobilize violence and outrage on social media. Also, psychologists have frequently used crime vignettes to test the effects of perceived violations on punitiveness, though this has been done mainly outside of a realistic news context (Fincher and Tetlock 2016; Tetlock et al. 2007).

To reduce the artificiality of our treatment, we constructed a news site that mimicked the Associated Press (AP) website and included a functioning comment section (see pp. 1-5 in the supplementary materials for a description of the website).[2] The use of the AP template is significant not only for its perceived objectivity but also because some extremist websites refer to it explicitly as a desirable format for presenting their own stories (Feinberg 2017). Additionally, the concise style of AP newswire articles makes the short form of a standard survey experiment vignette seem more realistic and nonpartisan. We had professional journalists at Politico and the New York Times vet our vignettes and our constructed website for plausibility and proper news formatting.

We used a factorial design that randomly assigned individuals to a combination of three two-level factor treatments, for a total of eight possible combinations. The treatment vector was a news article that described an alleged homicide, the identity of the suspect, and retaliatory violence against the suspect by a group of locals in a small American town (see Figures S2-S5, pp. 3-5 in the supplementary materials). The three treatment factors were: 1) the description of the homicide, sensationalized ("gruesome slaying of a local child") or

[2]The interactive survey web app was custom coded using the Python Django web framework and deployed to the web using the Amazon Elastic Beanstalk service. It was designed to be responsive, working on any screen size or device. Click data, comments, and browser metadata were stored in a PostgreSQL database hosted on Amazon Relational Database Services (RDS).

objective ("homicide of a local child"); 2) the presence of an outgroup cue, identifying the perpetrator as a "middle-aged male Muslim refugee" or not ("middle-aged male"); and 3) the content of the most upvoted comment (violent or conciliatory). Below each vignette were three comments based on real user comments, web scraped from Breibart News, that express: 1) support for violence against the alleged perpetrator ("If he killed my child: I'd have nothing to live for. I'd rain fiery retribution down on anyone who killed my child. His end would be brutal."); 2) neutral information-seeking about the event ("What could have driven this man to kill a child?"); and 3) conciliatory ("I hope that this conflict can be resolved peacefully. My heart goes out to the victims"). Regardless of the treatment condition, the top comment had 53 likes, the middle comment had eight, and the bottom comment had two. This distribution of 63 likes follows a power-law distribution that typically characterizes the distribution of likes on social media. In order to reduce experimenter demand, we included distractor questions before and after the treatment article that asked respondents about their online shopping habits.

The pre-treatment survey contained a battery of background questions concerning respondents' demographic characteristics, party identification, and individual dispositions towards authoritarianism, symbolic racism, and ethnocentrism. We use Feldman and Stenner (1997)'s four-item Child Rearing Values (CRV) scale to measure authoritarian personality. Our symbolic racism and ethnocentrism scales are taken directly from ANES. We calculate ethnocentrism by averaging outgroup feeling thermometers towards Asians, Blacks, Hispanics, Muslims, and refugees. We calculate symbolic racism by averaging the four items from the ANES symbolic racism scale, reversing two items to make the scale range from low to high racism.

The post-treatment survey asked a series of questions regarding respondents' emotional responses to the homicide and the attack on the alleged perpetrator, separately. Each emotion barometer asks respondents to indicate their felt intensity of eight emotions using five-point scales. To avoid biasing respondents' emotional reactions, we included a list of eight

emotions, positive and negative, based loosely on Robert Plutchik's typology of basic discrete emotions. We measured violent attitudes, directly and indirectly, using survey scale and open-ended responses. To measure punitiveness toward the alleged perpetrator in the story, we asked respondents to rate their support for nine punishments on seven-point (-3 to 3) scales. To test for support for extrajudicial violence, we ask if the perpetrator should be tortured and if he should be jailed without trial. Because direct questions about violence are subject to social desirability bias, we included indirect measures of punitiveness. We asked respondents how they think the locals who engaged in street violence against the suspect should be punished for beating the homicide suspect, and we ask if they would be willing to donate to a legal defense fund for the attackers. To measure attitudes toward outgroups, we presented respondents with four seven-point (-3 to 3) scales to indicate their support for or opposition to policies on refugee immigration, Muslim immigration, and monitoring and registering Muslim communities.

In addition to these survey scale responses, we collected behavioral data on respondents' interaction with the news website and text data from respondents' written reactions to the article. Our survey instructions encouraged respondents both to read the article and to interact with the website by liking, reporting, or posting comments in the comments section below the text of the news story, though doing so was not mandatory. Respondents' comments and any clicks on "report" or "like" buttons were stored in our database as behavioral measures of support or disapproval of violence. We asked respondents to provide twenty-word responses to the homicide and the mob attack on the suspect in two separate open-ended questions. We devised a novel typology to classify violent content (see Figure S14, p. 23), which a team of four researchers, blind to treatment condition, used to hand annotate all text data collected in open-ended responses and comments. Coding rules were pre-registered before our survey and involved categories (and subcategories) such as violence (lethal violence, extrajudicial violence, judicial violence, forcible physical displacement, property destruction/confiscation), group mentions (ingroup mention, race, religion,

etc.), and attacks (dehumanizing, demonizing). We detailed these coding rules in flow-charts that were used in weekly training sessions and as a reference for coders. In addition to these pre-registered categories, we coded responses according to emergent themes identified in treatment-blind reading of the comments[3] to help guide interpretation of our open-ended and survey scale responses. Agreement between coders was nearly perfect according to Cohen's kappa and F1 measures (see Table S15, p. 24).[4] After exiting the survey, we debrief respondents to ensure that they understood that the news article we presented was fictitious and why the study required this use of deception. The IRB reviewed this study and deemed it exempt; the research design, text coding scheme, and experimental vignettes for this study were pre-registered before conducting the study.

Our sample of 1655 respondents, recruited through Amazon's Mechanical Turk (MTurk) in early December 2018. To mitigate potential problems arising from automation and lack of demographic representativeness in MTurk samples, we collected extra metadata from respondents to identify bots, and purposively oversampled women, baby boomers, non-whites, and Republicans using Turk Prime. Despite our concerns with MTurk, many recent studies have suggested that MTurk samples can in fact be more reliable than traditional survey pools.[5] Our resulting sample roughly approximated the demographics of the American population. Within our sample, 49 percent identified as female and 77 percent as white, in line with the US Census Bureau's 2018 population estimates of 51 percent and 77 percent, respectively. The median age of our respondents was 36, versus 38 in the population. Looking at

[3]These categories included: *cognitive dissonance* between support for violence and the acknowledgement that violence is immoral, *information-seeking*, suspicion of *fake news*, suspicion of *racism*, and expressions of concerns for *social desirability* of one's true opinions on the article. See the appendix for coding diagrams and reliability statistics for these categories.

[4]To measure inter-coder reliability, two coders coded a random sample of open-ended responses. Cohen's kappa is an inter-coder reliability measure for nominal scales and two coders. Though there is no agreed-upon interpretation of the kappa coefficient, it has been suggested that 0.61-0.80 indicates substantial agreement, and 0.81-1.0 indicates near-perfect agreement. F1 is a common performance measure in machine learning ranging from 0 to 1. It is defined as the harmonic mean of precision and recall. F1 is a good measure of coder agreement when choices are imbalanced. With imbalance in choices, percent agreement (accuracy) will overestimate coder agreement.

[5]Recent work has found MTurk workers more reliable than subject pool workers in their level of attention (White et al. 2018) and when it comes to the issue of researcher demand (Hauser and Schwarz 2016).

Gallup's 2017 estimates of party affiliation, our sample was over-represented in self-identified Democrats (41.3 percent versus 29 percent) but proportionally representative of Republicans (28 percent). To ensure proportional geographic representativeness, we stratified our sampling according to the population size of each of the nine regional divisions designated by the US Census Bureau. This sample includes non-compliers who failed an attention check and respondents who received the treatment but did not finish the survey. Both noncompliance and attrition were extremely low: less than 1 percent failed the attention check, and 2.54 percent did not complete the entire post-treatment survey. Here we conduct an intent-to-treat analysis and do not exclude subjects on any post-treatment variables. The supplementary material (pp. 1-5) provides an in-depth description of the sampling and randomization process.

5 Results

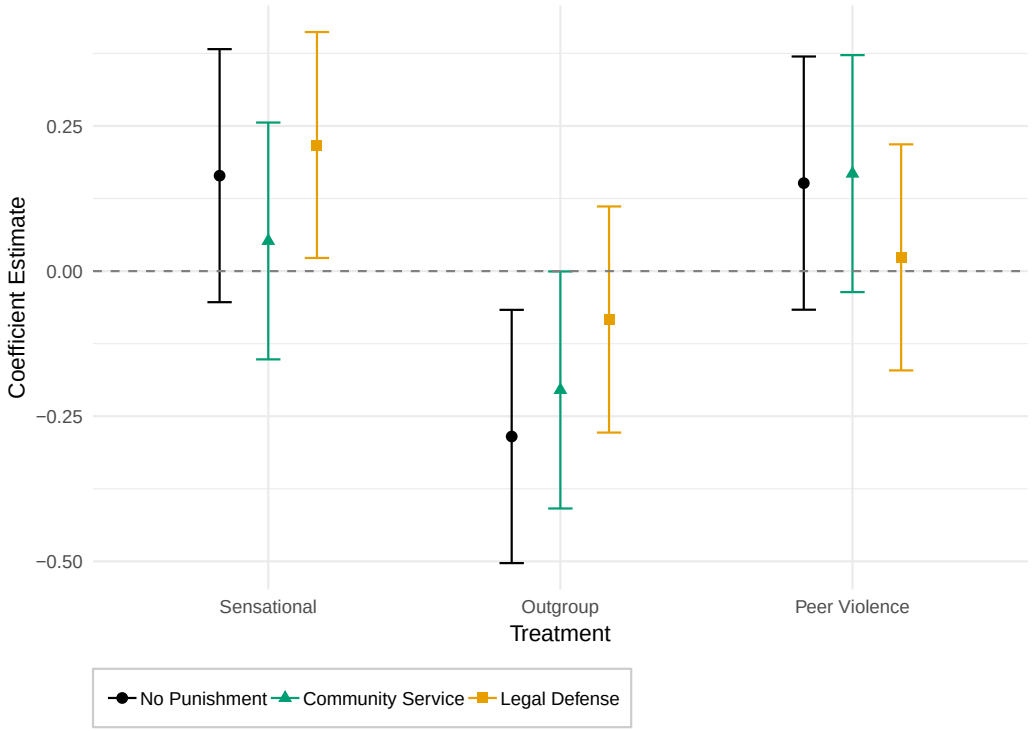
As expected for the sensational content hypothesis, the sensational content treatment significantly increased support for violence when measured indirectly through questions about the mob attack on the alleged perpetrator and open-ended responses on the events in the article.[6] Respondents presented with sensational content were more likely to donate to a legal defense fund for people who engaged in a mob attack on the alleged perpetrator ($P < 0.05$), and were more likely to believe that not punishing them was justified. However, this finding is not statistically different from zero at conventional levels of significance (Figure 1). This effect for sensational content further holds when looking at the text data from the open-ended responses. We find that sensational content significantly increased the probability that respondents will express support for violence in their written responses, including extrajudicial, mob-style violence ($P < 0.05$); for both measures this effect represents a roughly fifteen

[6]With the exception of support for detainment without trial, there are no significant average treatment effects for sensational content on punitiveness when looking at survey scale questions that directly asked respondents what kinds of punishments they would support for the alleged perpetrator in the article (see Figure S7, p. 7). We expected these direct measures to be vulnerable to social desirability bias.

percent increase in probability (Figure 2).[7]

A cursory comparison of representative open-ended responses illustrate differences in punitiveness across treatment conditions. One respondent given the sensationalized content treatment wrote the following, which was coded as supportive of violence: "I am absolutely disgusted, appalled, and at a loss for words about this attack. They should kill him in the street. Let everyone who wants a piece of him have a piece of him." Contrast this with a non-violent response in the control condition: "I want to know what happened. I want to know if the person that was beat up is the one accused of the murder. I want to know the details." See the supplementary material (pp. 16-17) for representative open-ended responses and comments by coding category and treatment condition.

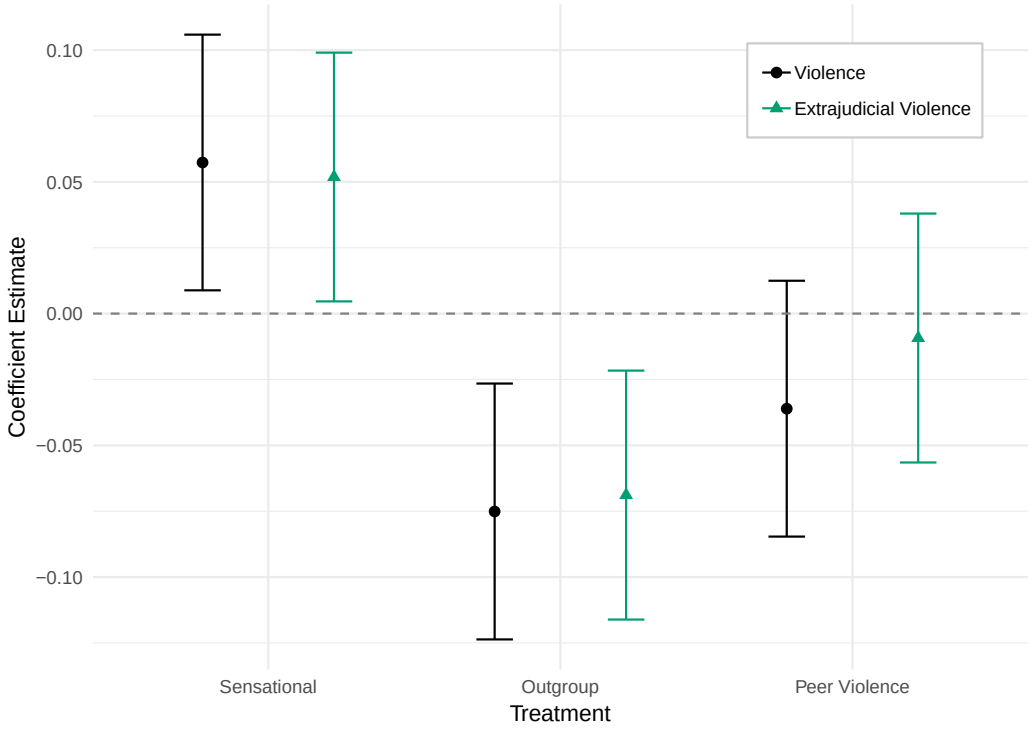
Figure 1: Support for the Mob Attack by Treatment Condition



Unexpectedly, outgroup cues decreased support for violence against the perpetrator in

[7]The survey scale and text measures of support were highly correlated. Expression of support for violence in the open-ended responses significantly and positively correlates with indirect survey scale responses as well as support for the death penalty for the alleged perpetrator (all $P < 0.001$) (see Figure S6 and Table S1, p. 6).

Figure 2: Support for Violence in the Open-ended Responses by Treatment Condition



the story when measured indirectly and in the open-ended responses ($P < 0.05$) and do not support the outgroup cue hypothesis (Figures 1 and 2). That being said, we do suspect this negative effect was driven by perceptions that the event or even the article itself was racist or fabricated. Respondents were significantly more likely to believe that the mob attack or the article itself was racially-motivated when given the outgroup treatment (see Figures S12-13, p. 22).

Our results support the bandwagoning hypothesis and shed light on how sensational content and violent peer influence affect interactions with online content. The sensational content treatment increased the probability that respondents would leave a comment on the website by roughly sixteen percent ($P < 0.05$) (Figure 3). Since the act of leaving a comment is vulnerable to significant selection effects, we run Heckman two-step selection models to detect and correct for selection bias (Heckman 1979). Older respondents and respondents scoring high on symbolic racism were far more likely to leave a comment. After taking this into account, we find that the sensational treatment increased the probability of writing

a violent comment by roughly twenty-four percent, though this effect is not statistically significant at conventional levels ($P = 0.14$ level) (Figure 4).[8]

While peer signals mattered little for attitudinal outcomes, they consistently affected how respondents interacted with the news article, creating a "cascade effect" (Bikhchandani, Hirshleifer and Welch 1992) for endorsement of violence. That is, respondents bandwagoned onto violent comments if they were already highly-liked. If the top, most-liked comment was violent, respondents were more likely to like a violent comment ($P < 0.05$) and to leave a violent comment of their own ($P < 0.05$); and far less likely to like a non-violent comment ($P < 0.001$) (Figures 3 and 4). Also, respondents were less likely to report the violent comment when it was highly liked, though this effect was not statistically significant ($P = 0.12$). Because the most liked comment was also positioned at the top of the three comments—as is usually the case in a comments section on a news website—this effect may possibly be driven by the position of the comment rather than the fact that it was highly liked; that is, respondents may have been merely satisficing by interacting with the upvoted violent comment because it was at the top of comment list. We believe the violent comment's upvoting was more important than its positioning for two reasons. First, we found that the upvoted violent comment treatment had a *negative* effect on reporting the violent comment and a *positive* treatment effect on expressing violence in a written comment, which suggests the effect is not due to merely interacting with whatever comment was positioned at the top of the comments list. Second, each comment was only one or two lines long, so users would have seen all three comments at once when they scrolled down; it is unlikely that the bottom two comments would have been less visible to respondents.

In line with the outrage-agression hypothesis, sensational content elicited strong emotional responses of anger ($P < 0.1$), fear ($P < 0.05$), and anxiety ($P < 0.05$) from respondents, while outgroup cues and peer support did not (Figure 5). To gauge whether emotional mechanisms mediated the treatment effect of sensational content on violent attitudes and

[8]We include effects that are not statistically significant because they offer suggestive evidence of our theory.

Figure 3: Behavioral Outcomes by Treatment Condition

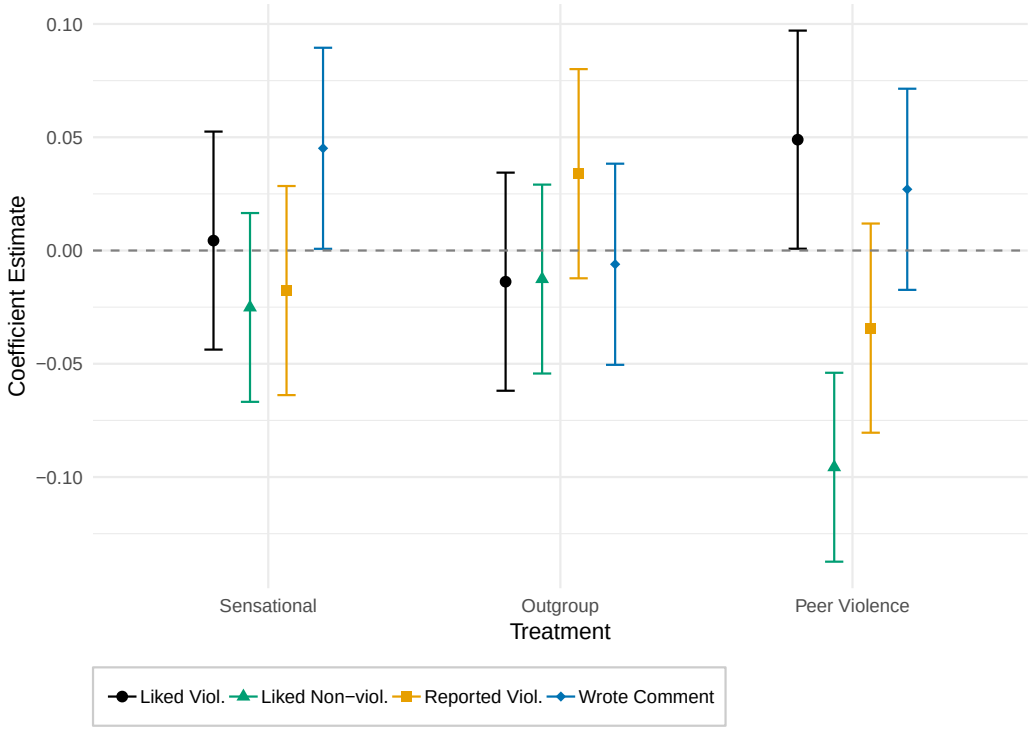
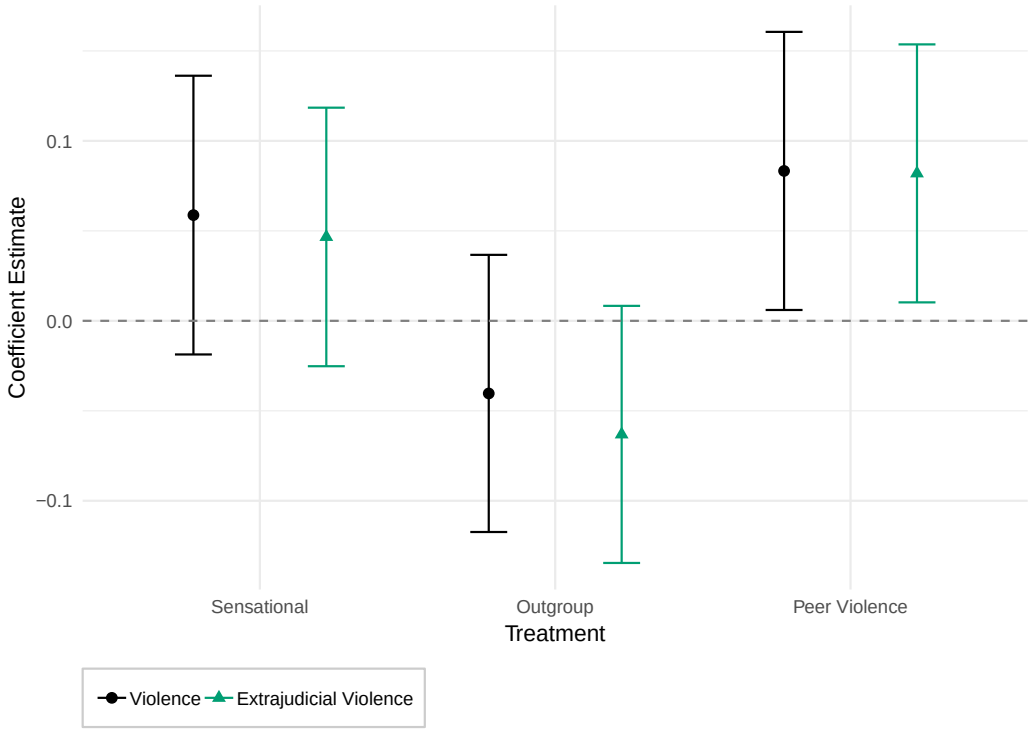
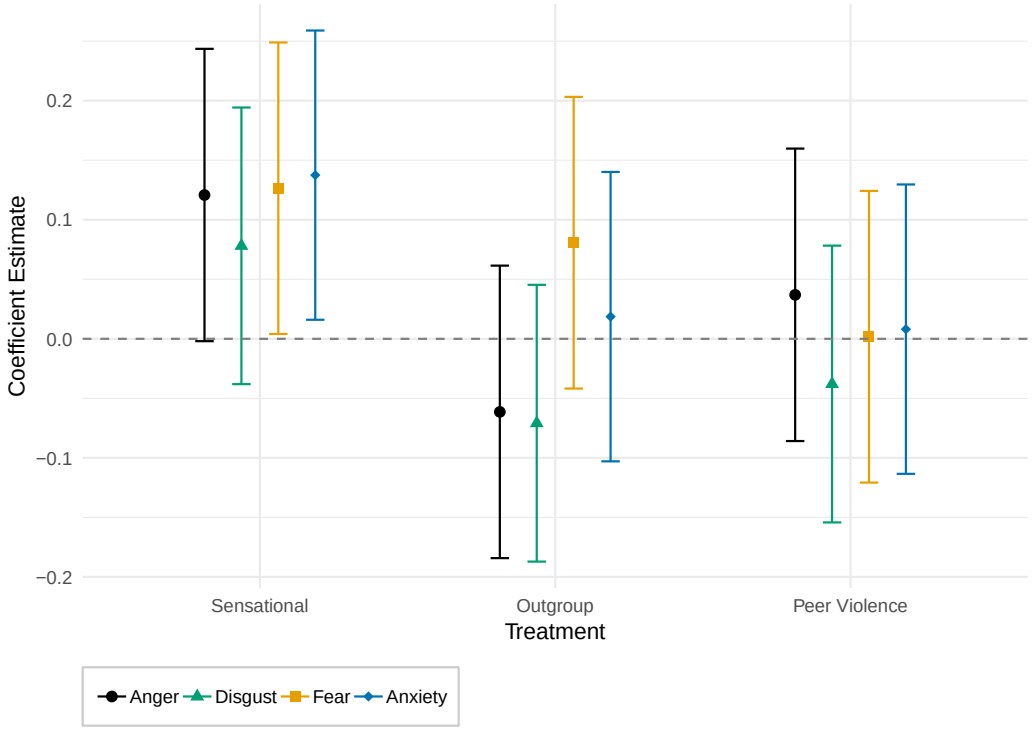


Figure 4: Comment Outcomes by Treatment Condition



behavior, we use causal mediation analysis (Imai, Jo and Stuart 2011). Because emotion was not randomly assigned, violating the assumption of sequential ignorability, we include possible confounding covariates to the models—i.e., measures of authoritarianism (Stenner 2005), ethnocentrism (Kinder and Kam 2010), symbolic racism (Banks 2014), and whether the respondent resided in the American South (Nisbett 1996). Causal mediation analysis revealed that anger, fear, and anxiety mediated the effect of sensational content on support for the mob attackers' legal defense fund ($ACME$ = 0.06, 0.04, and 0.03, respectively; all $P < 0.05$) and the likelihood of expressing support for violence in the open-ended responses ($ACME$ = 0.007, 0.003, and 0.002, respectively; all $P < 0.1$) (see Figures S8-10, pp. 13-14). We do not find emotional mediating effects for the other two treatments or behavioral and anti-outgroup outcomes.

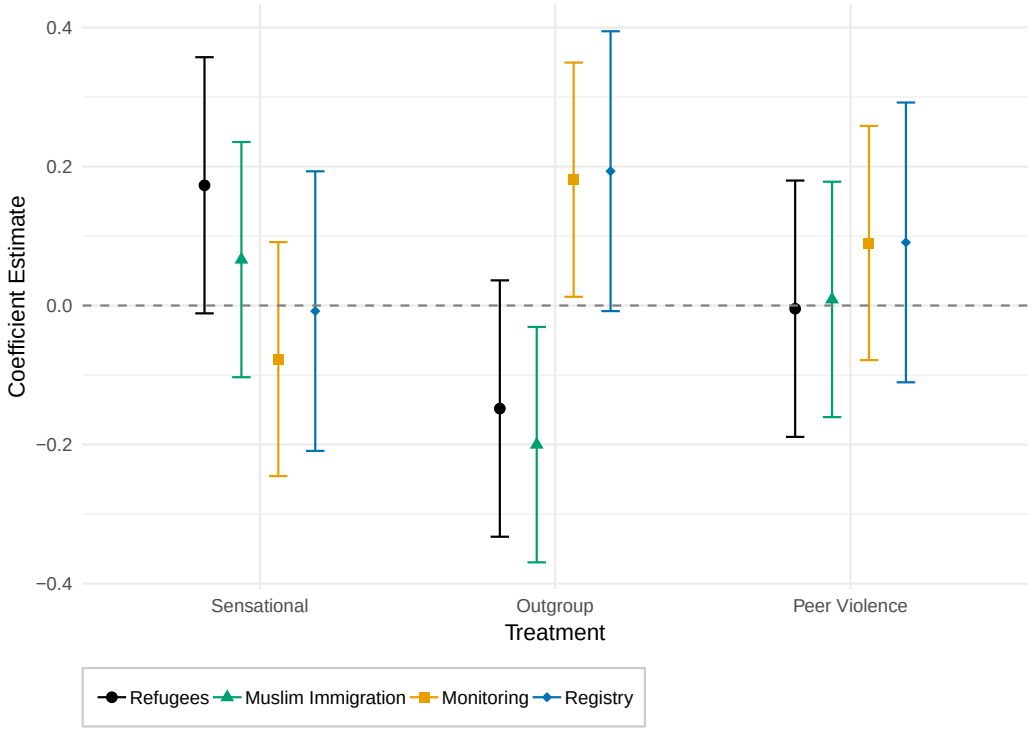
Figure 5: Emotional Response by Treatment Condition



Did this increased punitiveness translate into support for sanctions on Muslims and refugees? We find sensational content and peer support for violence did not increase support for anti-Muslim and anti-refugee policies; however, the mere inclusion of an outgroup cue on

average increased support for anti-Muslim and anti-refugee policies (Figure 6). Identifying the alleged perpetrator as a Muslim refugee caused respondents to support tightening refugee quotas ($P < 0.1$) and restricting Muslim immigration ($P < 0.05$), and positive for support for increased monitoring of Muslims ($P < 0.05$) and approval of a Muslim "registry" ($P < 0.1$). This finding contrasts with the earlier finding that including an outgroup cue increased individuals' suspicion that the attack described in the article, or the article itself, was racist. Indeed, this contradiction appeared in some of the open-ended responses in the outgroup cue treatment condition. One respondent in the outgroup cue treatment condition wrote: "I was thinking it was a fake story to invoke Islamophobia...I also do not want them moving next door to me or in our country, they can keep their backwards traditions in their own countries." Perceived racial bias or untrustworthiness of the content's source did not temper respondents' attitudes towards Muslims and refugees.

Figure 6: Support for Anti-Muslim Policies by Treatment Condition



This increased punitiveness towards refugees and Muslims may be a function of a general increase in punitiveness. Outrage arising from morally transgressive behavior may trigger

an "intuitive prosecutor" mindset that heightens one's desire to punish all subsequent transgressors, regardless of their connection to the original transgressor (Crockett 2017; Goldberg, Lerner and Tetlock 1999; Tetlock et al. 2007). To this end, we looked at the effects of the treatments on preferences regarding punishment of homicide, the crime featured in the article. We did not find a significant effect between sensational content, outgroup cues, or peer support on respondents' attitudes towards the punishment of homicide.

Contrary to our expectations, we did not find a significant interaction effect between outgroup cues and moral-emotional content on anti-immigrant and anti-Muslim policy preferences, and the effect is not in the expected direction. The average treatment effect for peer support is insignificant, as is its interaction between outgroup cues, though the signs of these effects are in their expected direction (see Table S5, p. 11).

6 Discussion

This study contributes to our understanding of sensational news and mass opinion in three ways. First, this is one of the first studies, to our knowledge, to demonstrate the causal effects of sensational news consumption on attitudes towards violence and online behavior; in contrast, the burgeoning literature on misinformation has focused mainly on its patterns of diffusion. Second, it identifies the importance of both anger and fear as emotional mediators of the effects of sensational news consumption on punitiveness. Third, we show that the threshold by which sensational news increases anti-outgroup sentiment is rather low: simply attributing a crime to a member of a certain outgroup is sufficient to increase discriminatory attitudes. Contrary to research that explicit outgroup cues do not increase support for anti-outgroup policies (Huber and Lapinski 2006; Mendelberg 2001) or are unnecessary (Banks 2016), we find that outgroup cues increase punitiveness towards an outgroup even when these cues raise concerns about racial bias.

Understanding the importance of online content in political life has perhaps never been

more critical than in the present. While these findings are a far cry from evidence that sensational news mobilizes on-the-ground violence, they underscore the importance of content and discourse in inflaming violent attitudes and carry significant policy implications for governments and social media companies working to stem the harmful effects of sensational news, especially sensationalized false news.

While outside the scope of this study, we believe that these findings are not restricted to the contemporary false news phenomenon or even online content: the use of sensational content to mobilize violence has a long genealogy. Why leaders choose these strategies for mobilizing popular support for violence and when they translate into on-the-ground violence remain crucial and promising areas for future scholarly inquiry.

References

Anderson, Adam K and Noam Sobel. 2003. "Dissociating intensity from valence as sensory inputs to emotion." *Neuron* 39(4):581–583.

Ansolabehere, Stephen and Shanto Iyengar. 1997. *Going negative: How political advertisements shrink and polarize the electorate*. The Free Press,.

Bandura, Albert, Bill Underwood and Michael E Fromson. 1975. "Disinhibition of aggression through diffusion of responsibility and dehumanization of victims." *Journal of research in personality* 9(4):253–269.

Banks, Antoine J. 2014. *Anger and racial politics: The emotional foundation of racial attitudes in America*. Cambridge University Press.

Banks, Antoine J. 2016. "Are group cues necessary? How anger makes ethnocentrism among whites a stronger predictor of racial and immigration policy opinions." *Political Behavior* 38(3):635–657.

Barrett, Devlin. 2018. "Hate crimes rose 17 percent last year, according to new FBI data." *Washington Post* .
URL: *https://web.archive.org/web/20190430073151/https://www.washingtonpost.com/world/national-security/hate-crimes-rose-17-percent-last-year-according-to-new-fbi-data/2018/11/13/e0dcf13e-e754-11e8-b8dc-66cca409c180_story.html?utm_term=.03f3360af144*

Baumeister, Roy F. 1999. *Evil: Inside human violence and cruelty*. Macmillan.

Beck, Aaron T. 1999. *Prisoners of hate: The cognitive basis of anger, hostility, and violence.* HarperCollins Publishers.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch. 1992. "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of political Economy* 100(5):992–1026.

Brader, Ted. 2005. "Striking a responsive chord: How political ads motivate and persuade voters by appealing to emotions." *American Journal of Political Science* 49(2):388–405.

Brader, Ted. 2006. *Campaigning for hearts and minds: How emotional appeals in political ads work*. University of Chicago Press.

Brady, William J, Julian A Wills, John T Jost, Joshua A Tucker and Jay J Van Bavel. 2017. "Emotion shapes the diffusion of moralized content in social networks." *Proceedings of the National Academy of Sciences* 114(28):7313–7318.

Charny, Israel W. 2019. *How Can We Commit the Unthinkable?: Genocide: the Human Cancer*. Routledge.

Charny, Israel W et al. 1982. "How can we commit the unthinkable." *Genocide: The Human Cancer (Boulder, CO: Westview Press, 1982)* 207.

Cohn, Norman. 1967. *Warrant for Genocide: The Myth of the Jewish World Conspiracy and the*. New York: Harper and Row.

Crockett, MJ. 2017. "Moral outrage in the digital age." *Nature Human Behaviour* 1(11):769.

Dower, John. 1986. *War without mercy: Race and power in the Pacific War*. Pantheon.

Drabman, Ronald S and Margaret H Thomas. 1974. "Does media violence increase children's toleration of real-life aggression?" *Developmental Psychology* 10(3):418.

Fein, Helen. 1979. *Accounting for genocide: National responses and Jewish victimization during the Holocaust*. New York: Free Press.

Feinberg, Ashley. 2017. "This is the daily stormer's playbook." *Huffington Post* .

Feldman, Stanley and Karen Stenner. 1997. "Perceived threat and authoritarianism." *Political Psychology* 18(4):741–770.

Fessler, DM and Kevin J Haley. 2003. "The strategy of affect: Emotions in human cooperation 12." *The Genetic and Cultural Evolution of Cooperation, P. Hammerstein, ed* pp. 7–36.

Fincher, Katrina M and Philip E Tetlock. 2016. "Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement." *Journal of Experimental Psychology: General* 145(2):131.

Fiske, Alan Page and Tage Shakti Rai. 2014. *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships*. Cambridge University Press.

Freedman, Paul and Ken Goldstein. 1999. "Measuring media exposure and the effects of negative campaign ads." *American journal of political Science* pp. 1189–1208.

Fujii, Lee Ann. 2004. "Transforming the moral landscape: the diffusion of a genocidal norm in Rwanda." *Journal of Genocide Research* 6(1):99–114.

Gates Jr, Henry Louis, Anthony P Griffin, Donald E Lively and Nadine Strossen. 1996. *Speaking of race, speaking of sex: Hate speech, civil rights, and civil liberties*. NYU Press.

Gilliam Jr, Franklin D and Shanto Iyengar. 2000. "Prime suspects: The influence of local television news on the viewing public." *American Journal of Political Science* pp. 560–573.

Gintis, Herbert, Eric Alden Smith and Samuel Bowles. 2001. "Costly signaling and cooperation." *Journal of theoretical biology* 213(1):103–119.

Goel, Vindu, Hari Kumar and Sheera Frenkel. 2018. "In Sri Lanka, Facebook Contends With Shutdown After Mob Violence." *New York Times* .
URL: *https://www.nytimes.com/2018/03/08/technology/sri-lanka-facebook-shutdown.html*

Goldberg, Julie H, Jennifer S Lerner and Philip E Tetlock. 1999. "Rage and reason: The psychology of the intuitive prosecutor." *European Journal of Social Psychology* 29(5-6):781–795.

Goldhagen, Daniel Jonah. 2009. *Worse than war: Genocide, eliminationism, and the ongoing assault on humanity*. Hachette UK.

Goldhagen, Daniel Jonah and Maurice Wohlgelernter. 1997. "Hitler's willing executioners." *Society* 34(2):32–37.

Hagan, John and Wenona Rymond-Richmond. 2008. *Darfur and the Crime of Genocide*. Cambridge University Press.

Hauser, David J. and Norbert Schwarz. 2016. "Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants." *Behavior Research Methods* 48(1):400–407.
URL: *https://doi.org/10.3758/s13428-015-0578-z*

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica: Journal of the econometric society* pp. 153–161.

Herf, Jeffrey. 2006. "The jewish enemy." *Nazi propaganda during World War II and the holocaust* .

Hern, Alex. 2018. "YouTube to Crack Down on Fake News, Backing 'Authoritative' Sources." *The Guardian* .
URL: *https://www.theguardian.com/technology/2018/jul/09/youtube-fake-news-changes*

Hill, Frances. 1995. *A delusion of Satan: The full story of the Salem witch trials*. Doubleday.

Huber, Gregory A and John S Lapinski. 2006. "The "race card" revisited: Assessing racial priming in policy contests." *American Journal of Political Science* 50(2):421–440.

Huber, Michaela, Leaf Van Boven, Bernadette Park and William T. Pizzi. 2015. "Seeing Red: Anger Increases How Much Republican Identification Predicts Partisan Attitudes and Perceived Polarization." *PLoS ONE* 10(9).

Huesmann, L Rowell and Laurie S Miller. 1994. Long-term effects of repeated exposure to media violence in childhood. In *Aggressive behavior*. Springer pp. 153–186.

Imai, Kosuke, Booil Jo and Elizabeth A Stuart. 2011. "Commentary: Using potential outcomes to understand causal mediation analysis." *Multivariate Behavioral Research* 46(5):861–873.

Kalmoe, Nathan P. 2014. "Fueling the fire: Violent metaphors, trait aggression, and support for political violence." *Political Communication* 31(4):545–563.

Kinder, Donald R and Cindy D Kam. 2010. *Us against them: Ethnocentric foundations of American opinion*. University of Chicago Press.

Kirkpatrick, Jennet. 2008. *Uncivil disobedience: Studies in violence and democratic politics*. Princeton University Press.

Kurzban, Robert, Peter DeScioli and Erin O'Brien. 2007. "Audience effects on moralistic punishment." *Evolution and Human behavior* 28(2):75–84.

Lazer, David MJ, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A Sloman, Cass R Sunstein, Emily A Thorson, Duncan J Watts and Jonathan L Zittrain. 2018. "The Science of Fake News." *Science* 359(6380):1094–1096.

Lerner, Jennifer S and Dacher Keltner. 2001. "Fear, anger, and risk." *Journal of Personality and Social Psychology* 81(1):146.

Lyons, Tessa. 2018. "Hard Questions: What's Facebook's Strategy for Stopping False News?" *Facebook Newsroom* .
URL: *https://newsroom.fb.com/news/2018/05/hard-questions-false-news/*

Marcus, George E, W Russell Neuman and Michael MacKuen. 2000. *Affective intelligence and political judgment*. University of Chicago Press.

Mendelberg, Tali. 2001. *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton University Press.

Mozur, Paul. 2018. "A Genocide Incited on Facebook, With Posts From Myanmar's Military." *New York Times* .
URL: *https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html*

Neubaum, G. and Nicole Kramer. 2017. *Monitoring the Opinion of the Crowd: Psychological Mechanisms Underlying Public Opinion Perceptions on Social Media*. Vol. 20.

Nisbett, Richard E. 1996. *Culture of honor: The psychology of violence in the South*. Routledge.

Noelle-Neumann, E. 1993. *The Spiral of Silence: Public Opinion, Our Social Skin*. University of Chicago Press.

Parth, M.N. and Shashank Bengali. 2018. "Rumors of Child-kidnapping Gangs and Other WhatsApp Hoaxes are Getting People Killed in India." *Los Angeles Times* .
URL: *https://www.latimes.com/world/asia/la-fg-india-whatsapp-2018-story.html*

Ryan, Timothy J. 2012. "What Makes Us Click?Demonstrating Incentives for Angry Discourse with Digital-Age Field Experiments." *Journal of Politics* 74(4):1138–1152.

Stenner, Karen. 2005. *The authoritarian dynamic*. Cambridge University Press.

Stets, Jan E. 2012. "Current emotion research in sociology: Advances in the discipline." *Emotion Review* 4(3):326–334.

Straus, Scott. 2007. "What is the relationship between hate radio and violence? Rethinking Rwanda's "radio machete"." *Politics & Society* 35(4):609–637.

Tetlock, Philip E., Penny S. Visser, Ramadhar Singh, Mark Polifroni, Amanda Scott, Sara Beth Elson, Philip Mazzocco and Phillip Rescober. 2007. "People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility." *Journal of Experimental Social Psychology* 43(2):195–209.

Tsesis, Alexander. 2002. *Destructive messages: How hate speech paves the way for harmful social movements*. Vol. 778 NYU Press.

Turner, Jonathan H and Jan E Stets. 2006. "Sociological theories of human emotions." *Annu. Rev. Sociol.* 32:25–52.

Valentino, Nicholas A, Ted Brader, Eric W Groenendyk, Krysha Gregorowicz and Vincent L Hutchings. 2011. "Election night's alright for fighting: The role of emotions in political participation." *The Journal of Politics* 73(1):156–170.

Valentino, Nicholas A, Vincent L Hutchings and Ismail K White. 2002. "Cues that matter: How political ads prime racial attitudes during campaigns." *American Political Science Review* 96(1):75–90.

Viterna, Jocelyn. 2013. *Women in war: The micro-processes of mobilization in El Salvador*. Oxford University Press.

Viterna, Jocelyn. 2014. "Radical or Righteous? Using Gender to Shape Public Perceptions of Political Violence." *Dynamics of Political Violence: A Process-Oriented Perspective on Radicalization and the Escalation of Political Conflict* pp. 189–216.

Vosoughi, Soroush, Deb Roy and Sinan Aral. 2018. "The spread of true and false news online." *Science* 359(6380):1146–1151.

Waldron, Jeremy. 2012. *The harm in hate speech*. Harvard University Press.

White, Ariel, Anton Strezhnev, Christopher Lucas, Dominika Kruszewska and Connor Huff. 2018. "Investigator Characteristics and Respondent Behavior in Online Surveys." *Journal of Experimental Political Science* 5(1):56–67.

Wood, Elisabeth Jean. 2003. *Insurgent collective action and civil war in El Salvador*. Cambridge University Press.

Yanagizawa-Drott, David. 2014. "Propaganda and conflict: Evidence from the Rwandan genocide." *The Quarterly Journal of Economics* 129(4):1947–1994.